Unraveling Spurious Properties of Interaction Networks with Tailored Random Networks

Stephan Bialonski,[1, 2, 3, *] Martin Wendler,[4] and Klaus Lehnertz[1, 2, 3]

[1]*Department of Epileptology, University of Bonn, Bonn, Germany*
[2]*Helmholtz Institute for Radiation and Nuclear Physics, University of Bonn, Bonn, Germany*
[3]*Interdisciplinary Center for Complex Systems, University of Bonn, Bonn, Germany*
[4]*Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany*

We investigate interaction networks that we derive from multivariate time series with methods frequently employed in diverse scientific fields such as biology, quantitative finance, physics, earth and climate sciences, and the neurosciences. Mimicking experimental situations, we generate time series with finite length and varying frequency content but from independent stochastic processes. Using the correlation coefficient and the maximum cross-correlation, we estimate interdependencies between these time series. With clustering coefficient and average shortest path length, we observe unweighted interaction networks, derived via thresholding the values of interdependence, to possess non-trivial topologies as compared to Erdős-Rényi networks, which would indicate small-world characteristics. These topologies reflect the mostly unavoidable finiteness of the data, which limits the reliability of typically used estimators of signal interdependence. We propose random networks that are tailored to the way interaction networks are derived from empirical data. Through an exemplary investigation of multichannel electroencephalographic recordings of epileptic seizures—known for their complex spatial and temporal dynamics—we show that such random networks help to distinguish network properties of interdependence structures related to seizure dynamics from those spuriously induced by the applied methods of analysis.

PACS numbers:

Introduction

The last years have seen an extraordinary success of network theory and its applications in diverse disciplines, ranging from sociology, biology, earth and climate sciences, quantitative finance, to physics and the neurosciences [1–4]. There is now growing evidence that research into the dynamics of complex systems profits from a network perspective. Within this framework, complex systems are considered to be composed of interacting subsystems. This view has been adopted in a large number of modeling studies and empirical studies. It is usually assumed that the complex system under study can be described by an *interaction network*, whose nodes represent subsystems and whose links represent interactions between them. Interaction networks derived from empirical data (multivariate time series) have been repeatedly studied in climate science (climate networks, see [5–9] and references therein), in seismology (earthquake networks, see, e.g., [10–13]), in quantitative finance (financial networks, see e.g. [14–18] and references therein), and in the neurosciences (brain functional networks, see [19, 20] for an overview). Many interaction networks have been reported to possess non-trivial properties such as small-world architectures, community structures, or hubs (nodes with high centrality), all of which have been considered to be characteristics of the dynamics of the complex system.

When analyzing empirical data one is faced with the challenge of defining nodes and inferring links from multivariate noisy time series with only a limited number of data points due to stationarity requirements. Different approaches varying to some degree across disciplines have been proposed. For most approaches, each single time series is associated with a node and inference of links is based on time series analysis techniques. Bivariate time series analysis methods, such as the correlation coefficient, are used as estimators of signal interdependence which is assumed to be indicative of an interaction between different subsystems. Inferring links from estimates of signal interdependence can be achieved in different ways. Weighted interaction networks can be derived by considering estimated values of signal interdependence (sometimes mapped via some function) as link weights. Since methods characterizing unweighted networks are well-established and readily available, such networks are more frequently derived from empirical data. Besides approaches based on constructing minimum spanning trees (see, e.g., reference [14]), on significance testing [21–23], or on rank-ordered network growth (see, e.g., reference [15]), a common practice pursued in many disciplines is to choose a threshold above which an estimated value of signal interdependence is converted into a link ("thresholding", see, e.g., references [5, 12, 16, 20]). Following this approach, the resulting unweighted interaction networks have been repeatedly investigated employing various networks characteristics, among which we mention the widely-used clustering coefficient C and average shortest path length L to assess a potential small-world characteristic, and the node degrees in order to identify

*Electronic address: bialonski@gmx.net

hubs.

As studies employing the network approach grow in numbers, the question arises as to how informative reported results are with respect to the investigated dynamical systems. To address this issue, properties of interaction networks are typically compared to those obtained from network null models. Most frequently, Erdős-Rényi random networks [24] or random networks with a predefined degree distribution [25, 26] serve as null models; network properties that deviate from those obtained from the null model are considered to be characteristic of the investigated dynamical system. Only in a few recent studies, results obtained from network analyses have been questioned in relation to various assumptions underlying the network analysis approach. Problems pointed out include: incomplete data sets and observational errors in animal social network studies [27]; representation issues and questionable use of statistics in biological networks (see [28] and references therein); challenging node and link identification in the neurosciences [29–31]; the issue of spatial sampling of complex systems [31–33]. This calls not only for a careful interpretation of results but also for the development of appropriate null models that incorporate knowledge about the way networks are derived from empirical data.

We study – from the perspective of field data analysis – a fundamental assumption underlying the network approach, namely that the multivariate time series are obtained from interacting dynamical processes and are thus well represented by a model of mutual relationships (i.e., an interaction network). Visual inspection of empirical time series typically reveals a perplexing variety of characteristics ranging from fluctuations on different time scales to quasi-periodicity suggestive of different types of dynamics. Moreover, empirical time series are inevitably noisy and finite leading to a limited reliability of estimators of signal interdependencies. This is aggravated with the advent of time-resolved network analyses where a good temporal resolution often comes at the cost of diminished statistics. Taken together, it is not surprising that the suitability of the network approach is notoriously difficult to judge prior to analysis.

We here employ the above-mentioned thresholding-approach to construct interaction networks for which we estimate signal interdependence with the frequently used correlation coefficient and the maximum cross correlation. We derive these networks, however, from multivariate time series of finite length that are generated by independent (non-interacting) processes which would a priori not advocate the notion of a representation by a model of mutual relationships. In simulation studies we investigate often used network properties (clustering coefficient, average shortest path length, number of connected components). We observe that network properties can deviate pronouncedly from those observed in Erdős-Rényi networks depending on the length and the spectral content of the multivariate time series. We address the question whether similar dependencies can also be observed in empirical data by investigating multichannel electroencephalographic (EEG) recordings of epileptic seizures that are known for their complex spatial and temporal dynamics. Finally, we propose random networks that are tailored to the way interaction networks are derived from multivariate empirical time series.

Methods

Interaction networks are typically derived from N multivariate time series x_i ($i \in \{1, \ldots, N\}$) in two steps. First, by employing some bivariate time series analysis method, interdependence between two time series $x_i(t)$ and $x_j(t)$ ($t \in \{1, \ldots, T\}$) is estimated as an indicator for the strength of interaction between the underlying systems. A multitude of estimators [34–40], which differ in concepts, robustness (e.g., against noise contaminations), and statistical efficiency (i.e., the amount of data required), is available. Studies that aim at deriving interaction networks from field data frequently employ the absolute value of the linear correlation coefficient to estimate interdependence between two time series. The entries of the correlation matrix $\boldsymbol{\rho}^{\mathrm{c}}$ then read

$$\rho_{ij}^{\mathrm{c}} := \left| T^{-1} \sum_{t=1}^{T} (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)\hat{\sigma}_i^{-1}\hat{\sigma}_j^{-1} \right| \quad (1)$$
$$=: \left| \mathrm{corr}(x_i, x_j) \right|,$$

where $\bar{x}_i$ and $\hat{\sigma}_i$ denote mean value and the estimated standard deviation of time series x_i. Another well established method to characterize interdependencies is the cross correlation function. Here we use the maximum value of the absolute cross correlation between two time series,

$$\rho_{ij}^{\mathrm{m}} := \max_{\tau} \left\{ \left| \frac{\xi(x_i, x_j)(\tau)}{\sqrt{\xi(x_i, x_i)(0)\xi(x_j, x_j)(0)}} \right| \right\}, \quad (2)$$

with

$$\xi(x_i, x_j)(\tau) := \begin{cases} \sum_{t=1}^{T-\tau} x_i(t+\tau)x_j(t) & , \tau \geq 0 \\ \xi(x_j, x_i)(-\tau) & , \tau < 0 \end{cases} \quad (3)$$

to define the entries of the cross correlation matrix $\boldsymbol{\rho}^{\mathrm{m}}$. As practiced in field data analysis, we normalize the time series to zero mean before pursuing subsequent steps of analysis. Note that ρ_{ij}^{m} is then the maximum value of the absolute cross covariance function. Both interdependence estimators are symmetric ($\rho_{ij}^{\mathrm{c}} = \rho_{ji}^{\mathrm{c}}$ and $\rho_{ij}^{\mathrm{m}} = \rho_{ji}^{\mathrm{m}}$) and are confined to the interval $[0, 1]$. High values indicate strongly interdependent time series while dissimilar time series result in values close to zero for T sufficiently large.

Second, the adjacency matrix $\mathbf{A}$ representing an unweighted undirected interaction network is derived from $\boldsymbol{\rho}^{\mathrm{c}}$ (or $\boldsymbol{\rho}^{\mathrm{m}}$) by thresholding. For a threshold $\theta \in [0, 1]$

entries A_{ij} and A_{ji} of $\mathbf{A}$ are set to 1 (representing an undirected link between nodes i and j) for all entries $\rho_{ij}^{\mathrm{c}} > \theta$ ($\rho_{ij}^{\mathrm{m}} > \theta$, respectively) with $i \neq j$, and to zero (no link) otherwise. In many studies θ is not chosen directly but determined such that the derived network possesses a previously specified mean degree $\bar{k} := N^{-1} \sum_i k_i$, where k_i denotes the degree of i, i.e., the number of links connected to node i. More frequently, θ is chosen such that the network possesses a previously specified link density $\epsilon = \bar{k}(N-1)^{-1}$. We will follow the latter approach and derive networks for fixed values of ϵ.

To characterize a network as defined by $\mathbf{A}$, a plethora of methods have been developed. Among them, the clustering coefficient C and the average shortest path length L are frequently used in field studies. The local clustering coefficient C_i is defined as

$$C_i := \begin{cases} \frac{1}{k_i(k_i-1)} \sum_{j,m} A_{ij} A_{jm} A_{mi}, & \text{if } k_i > 1 \\ 0, & \text{if } k_i \in \{0, 1\}. \end{cases} \quad (4)$$

C_i represents the fraction of the number of existing links between neighbors of node i among all possible links between these neighbors [1, 2, 41]. The clustering coefficient C of the network is defined as the mean of the local clustering coefficients,

$$C := \frac{1}{N} \sum_{i=1}^{N} C_i. \quad (5)$$

C quantifies the local interconnectedness of the network and $C_i, C \in [0, 1]$.

The average shortest path length is defined as the average shortest distance between any two nodes,

$$\tilde{L} := \frac{2}{N(N+1)} \sum_{i \leq j} l_{ij}, \quad (6)$$

and characterizes the overall connectedness of the network. l_{ij} denotes the length of the shortest path between nodes i and j. The definition of the average shortest path length varies across the literature. Like some authors, we here include the distance from each node to itself in the average ($l_{ii} = 0 \forall i$). Exclusion will, however, just change the value by a constant factor of $(N+1)/(N-1)$.

If a network disintegrates into a number N_c of different connected components, there will be pairs of nodes (i, j), for which no connecting path exists, in which case one usually sets $l_{ij} = \infty$ and thus $\tilde{L} = \infty$. In order to avoid this situation, in some studies l_{ij} in equation (6) is replaced by l_{ij}^{-1}. The quantity defined this way is called efficiency [42, 43]. Another approach, which we will follow here and which is frequently used in field studies, is to exclude infinite values of l_{ij} from the average. The average shortest path length then reads

$$L := \frac{1}{|S|} \sum_{(i,j) \in S} l_{ij}, \quad (7)$$

where

$$S := \{(i, j) \mid l_{ij} < \infty; \; i, j = 1, \ldots, N\} \quad (8)$$

denotes the set of all pairs (i, j) of nodes with finite shortest path. The number of such pairs is given by $|S|$. Note that $L \to 0$ for $N_c \to N$.

In field studies, values of C and L obtained for interaction networks are typically compared with average values obtained from an ensemble of random Erdős-Rényi (ER) networks [24]. Between every pair of nodes is a link with probability ϵ, and links for different pairs exist independently from each other. The expectation value of the clustering coefficient of ER networks is $C_{\mathrm{ER}} = \epsilon$ [2]. The dependence of the average shortest path length L_{ER} of ER networks on ϵ and N is more complicated (see references [2, 44]). Almost all ER networks are connected, if $\epsilon \gg \ln N/(N-1)$. ER networks with a predefined number of links (and thus link density) can also be generated by successively adding links between randomly chosen pairs of nodes until the predefined number of links is reached. During this process, multiple links between nodes are avoided.

Results

Simulation studies

We consider time series z_i, $i \in \{1, \ldots, N\}$, whose entries $z_i(t)$ are drawn independently from the uniform probability distribution $\mathcal{U}$ on the interval $(0, 1)$. We will later study the impact of different lengths T of these random time series on network properties. In order to enable us to study the effects of different spectral contents on network properties, we add the possibility to low-pass filter z_i by considering

$$x_{i,M,T}(t) := M^{-1} \sum_{l=t}^{t+M-1} z_i(l), \qquad z_i(l) \sim \mathcal{U}, \quad (9)$$

where $t \in \{1, \ldots, T\}$, and $1 \leq M \ll T$. By definition $x_{i,1,T}(t) = z_i(t) \forall t$. With the size M of the moving average we control the spectral contents of time series. We here chose this ansatz for the sake of simplicity, for its mathematical treatability, and because the random time series with different spectral contents produced this way show all properties we want to illustrate.

In the following we will study the influence of the length T of time series on network properties by considering $x_{i,1,T}$ for different T. For a chosen value of T we determine R realizations of $x_{i,1,T}$ and we denote each realization r with $x_{i,1,T}^{(r)}$. When studying the influence of the spectral content we will consider $x_{i,M,T'}$ with different M and with $T' = 500$. We chose this value of T' because we are interested in investigating time series of short length as typically considered in field studies. For a

chosen value of M we determine R realizations of $x_{i,M,T'}$ and we denote realization r with $x_{i,M,T'}^{(r)}$.

In order to keep the line of reasoning short and clear, we will present supporting and more rigorous mathematical results in Appendix A and refer to them in places where needed. In addition, since we observed most simulation studies based on $\boldsymbol{\rho}^{\mathrm{m}}$ to yield qualitatively the same results as those based on $\boldsymbol{\rho}^{\mathrm{c}}$, we will present results based on $\boldsymbol{\rho}^{\mathrm{c}}$ only and report results of our studies based on $\boldsymbol{\rho}^{\mathrm{m}}$ whenever we observed qualitative differences.

Clustering coefficient

Let $\rho_{ij,1,T}^{(r)} := \rho^{\mathrm{c}}(x_{i,1,T}^{(r)}, x_{j,1,T}^{(r)})$ denote the absolute value of the empirical correlation coefficient estimated for time series $x_{i,1,T}^{(r)}$ and $x_{j,1,T}^{(r)}$, and let us consider R realizations, $r \in \{1, \ldots, R\}$. Because of the independence of processes generating the time series and because of the symmetry of the correlation coefficient, we expect the R values of the empirical correlation coefficient calculated for finite and fixed T to be distributed around the mean value 0. The variance of this distribution will be higher the lower we choose T. If we sample one value $\rho_{ij,1,T}^{(r)}$ out of the R values it is almost surely that $\rho_{ij,1,T}^{(r)} > 0$. Thus there are thresholds $0 < \theta < \rho_{ij,1,T}^{(r)}$ for which we would establish a link between the corresponding nodes i and j when deriving a network. Let us now consider a network of N nodes whose links are derived from N time series as before. For some $\theta > 0$ the network will possess links and $\epsilon > 0$. We expect to observe ϵ for some fixed $\theta > 0$ to be higher the larger the variance of the distribution of $\rho_{ij,1,T}^{(r)}$. Likewise, for fixed values of ϵ we expect to find θ to be higher the lower we choose a value of T.

As a first check of this intuition we derive an approximation ϵ_{al} for the edge density by making use of the asymptotic limit ($T \to \infty$, see Appendix A, Lemma 2 for details),

$$\epsilon_{\mathrm{al}}(\theta, T) = 2\Phi(-\sqrt{T}\theta), \tag{10}$$

where Φ denotes the cumulative distribution function of a standard normal distribution. In figure 1 (top left) we show the dependence of $\epsilon_{\mathrm{al}}(\theta, T)$ on θ for exemplary values of T. Indeed, $\epsilon_{\mathrm{al}}(\theta, T)$ is decreasing in θ and T.

The concession of taking the asymptotic limit when deriving equation (10) may limit its validity in the case of small values of T in which we are especially interested. Thus, we approach this case by simulation studies. Let us consider $R = 10^6$ values of $\rho_{12,M,T}^{(r)}$ obtained for R realizations of two time series $x_{i,M,T}$, $i \in \{1, 2\}, r \in \{1, \ldots, R\}$. We estimate the edge density $\hat{\epsilon}(\theta, M, T)$ by

$$\hat{\epsilon}(\theta, M, T) := R^{-1} \sum_r H_{12,M,T}^{(r)}(\theta), \tag{11}$$

where $H_{ij,M,T}^{(r)}(\theta) = 1$ for $\rho_{ij,M,T}^{(r)} > \theta$, and 0 else. Note that $\hat{\epsilon}(\theta, M, T)$ does not depend on N. This is because $\hat{\epsilon}(\theta, M, T)$ represents the (numerically determined) probability that there is a link between two vertices. The dependence of $\hat{\epsilon}(\theta, 1, T)$ on θ for different values of T shown in figure 1 (top left) indicates a good agreement between $\epsilon_{\mathrm{al}}(\theta, T)$ and $\hat{\epsilon}(\theta, 1, T)$ for larger values of T but an increasing difference for $T < 30$.

We proceed by estimating the clustering coefficient $\hat{C}$ for our model using R realizations of three time series $x_{i,M,T}$, $i \in \{1, \ldots, 3\}$ by

$$\hat{C}(\theta, M, T) := \frac{\sum_r H_{12,M,T}^{(r)}(\theta) H_{13,M,T}^{(r)}(\theta) H_{23,M,T}^{(r)}(\theta)}{\sum_r H_{12,M,T}^{(r)}(\theta) H_{13,M,T}^{(r)}(\theta)}. \tag{12}$$

The dependence of $\hat{C}(\theta, 1, T)$ on θ for various T is shown in the top right part of figure 1. For fixed T, $\hat{C}(\theta, 1, T)$ decreases from 1 with increasing values of θ which one might expect due to the decrease of ϵ. However, we also observe for $\theta > 0$ that $\hat{C}(\theta, 1, T)$ takes on higher values the lower T.

In order to investigate whether the clustering coefficients of our networks differ from those of Erdős-Rényi networks we use equation (11) and obtain $\hat{C}_{M,T}(\epsilon) = \hat{C}(\hat{\theta}(\epsilon, M, T), M, T)$ with $\hat{\theta}(\epsilon, M, T) = \inf\{\theta : \hat{\epsilon}(\theta, M, T) \geq \epsilon\}$. This allows the comparison with $C_{\mathrm{ER}}(\epsilon) = \epsilon$ by considering the ratio $\hat{\gamma}(\epsilon, M, T) := \hat{C}_{M,T}(\epsilon)/C_{\mathrm{ER}}(\epsilon)$. Remarkably, $\hat{\gamma}(\epsilon, 1, T) \gg 1$ for a range of values of ϵ and T (cf. lower left part of figure 1). $\hat{\gamma}(\epsilon, 1, T)$ even increases for small ϵ and T. This indicates that there is a relevant dependence between the three random variables $\rho_{ij,M,T}$, $\rho_{il,M,T}$, and $\rho_{jl,M,T}$ for different indices i, j, l and small T. For $T \to \infty$ and fixed edge density, C converges to C_{ER} because the dependence between the random variables $\rho_{ij,M,T}$, $i, j \in \{1, \ldots, N\}$, vanishes (i.e., the random variables will converge in distribution to independent normal random variables).

In order to gain deeper insights into the influence of the spectral contents of random time series on the clustering coefficient, we repeat the steps of analysis with time series $x_{i,M,T'}$, where $T' = 500$ is kept fix, and we choose different values of M. Figure 1 (top panels and lower left) shows that the higher the amount of low-frequency contributions (large M) the higher $\hat{\epsilon}(\theta, M, T')$ and $\hat{C}(\theta, M, T')$ (for $\theta > 0$), and the higher $\hat{\gamma}(\epsilon, M, T')$ (for $\epsilon \ll 1$). The difference between Erdős-Rényi networks and our time series networks becomes more pronounced ($\hat{\gamma}(\epsilon, M, T') \gg 1$) the smaller ϵ and the higher M.

Given the similar dependence of $\hat{\gamma}$, $\hat{C}$, and $\hat{\epsilon}$ on T and M, we hypothesize that the similarity can be traced back to similar variances of $\rho_{ij,1,T}$ and $\rho_{ij,M,T'}$ for some values of T and M. By making use of the asymptotic variance of the limit distributions of $T \to \infty$, we derive an expression relating $\mathrm{Var}(\rho_{ij,1,T})$ and $\mathrm{Var}(\rho_{ij,M,T})$ to each other (see
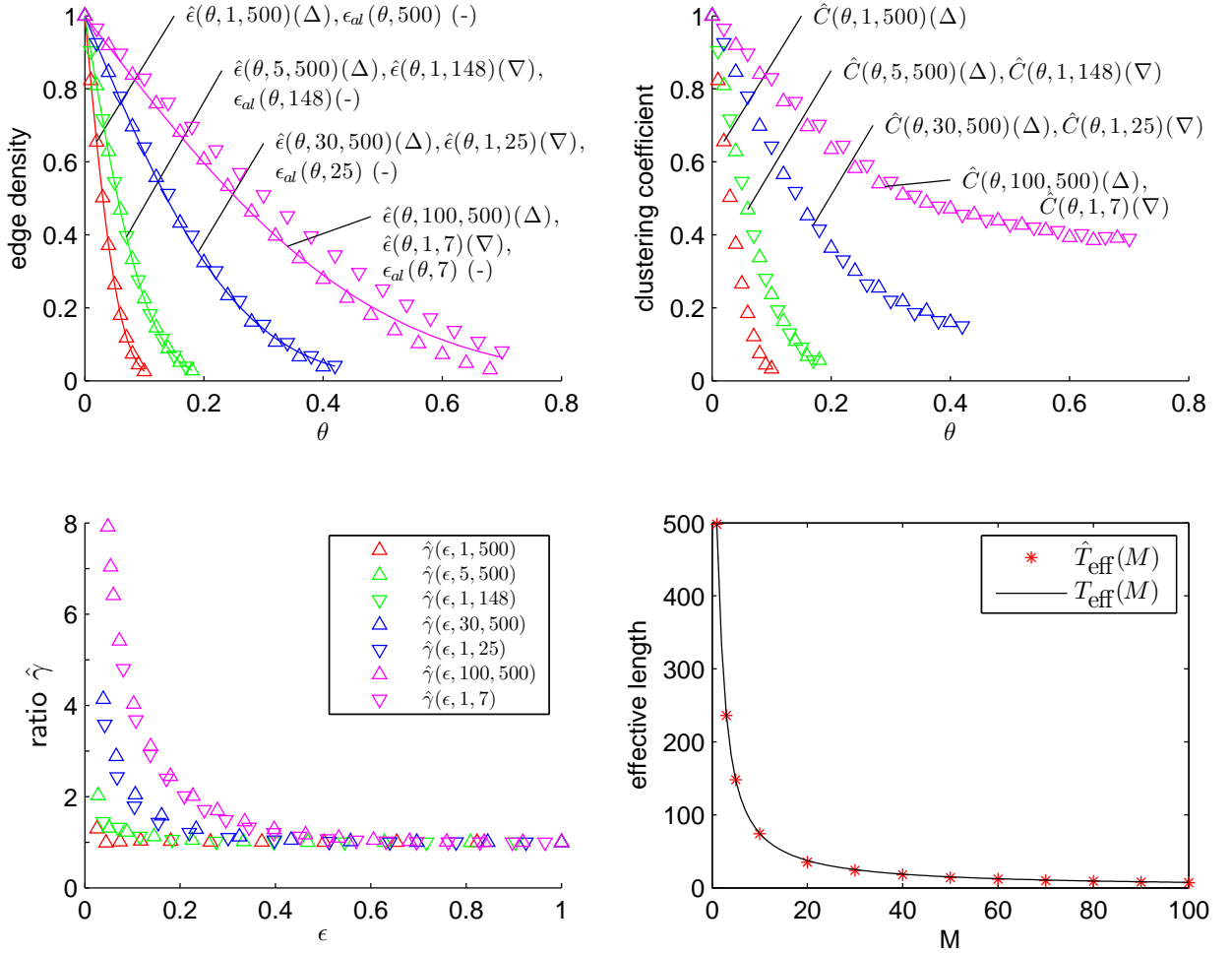
FIG. 1: **Simulation results for the edge density, the clustering coefficient, and the effective length.** Top row: Dependence of edge density $\hat{\epsilon}(\theta, M, T)$ (left) and of clustering coefficient $\hat{C}(\theta, M, T)$ (right) on the threshold θ for different values of the size M of the moving average and of the length T of time series. Values of edge density $\epsilon_{\text{al}}(\theta, T)$ obtained with the asymptotic limit (equation (10)) are shown as lines (top left). Bottom left: Dependence of the ratio $\hat{\gamma}(\epsilon, M, T) = \hat{C}_{M,T}(\epsilon)/C_{\text{ER}}(\epsilon)$ on edge density ϵ. Note, that we omitted values of estimated quantities obtained for $\theta \in \{\theta : (R^{-1}\sum_r H_{12,M,T}^{(r)}(\theta)H_{13,M,T}^{(r)}(\theta)) < 10^{-3}\}$ since the accuracy of the statistics is no longer guaranteed. Bottom right: Dependence of effective length T_{eff} as determined by equation (14) (black line) and its numerical estimate $\hat{T}_{\text{eff}}$ (red markers) on M.

Appendix A, Lemma 1),

$$\text{Var}(\rho_{ij,M,T}) \approx g(M)\text{Var}(\rho_{ij,1,T}),$$

$$\text{with } g(M) = \frac{2}{3}M + \frac{1}{3M}. \quad (13)$$

We are now able to define an effective length T_{eff} of time series,

$$T_{\text{eff}}(M) := \frac{T'}{g(M)}, \quad (14)$$

for which $\text{Var}(\rho_{ij,1,T_{\text{eff}}}) \approx \text{Var}(\rho_{ij,M,T'})$. In the lower right part of figure 1 we show $T_{\text{eff}}(M)$ in dependence on M. To investigate whether equation (14) also holds for small values of T, we determine numerically, for different values of θ, $\hat{C}(\theta, 1, T)$ for $T \in \{3, \ldots, T'\}$ as well

as $\hat{C}(\theta, M, T')$ for some chosen values of M. Eventually, we determine for each value of M a value of T, for which $\hat{C}(\theta, 1, T)$ and $\hat{C}(\theta, M, T')$ curves match in a least-squares sense, and denote this value with $\hat{T}_{\text{eff}}$ (see the lower right part of figure 1). We observe a maximum deviation $\left|T_{\text{eff}} - \hat{T}_{\text{eff}}\right| \approx 2$ and conclude that equation (14) indeed holds for small length T of time series. Moreover, numerically determined dependencies of $\hat{\epsilon}$ on θ, $\hat{C}$ on θ, as well as $\hat{\gamma}$ on ϵ for pairs of values (M, T') show a remarkable similarity to those dependencies obtained for pairs of values $(1, \hat{T}_{\text{eff}})$.

Thus, the clustering coefficient of networks derived from random time series of finite length and/or with a large amount of low-frequency contributions is higher than the one of Erdős-Rényi (ER) networks – independently of the network size N (cf. equation (12)). This
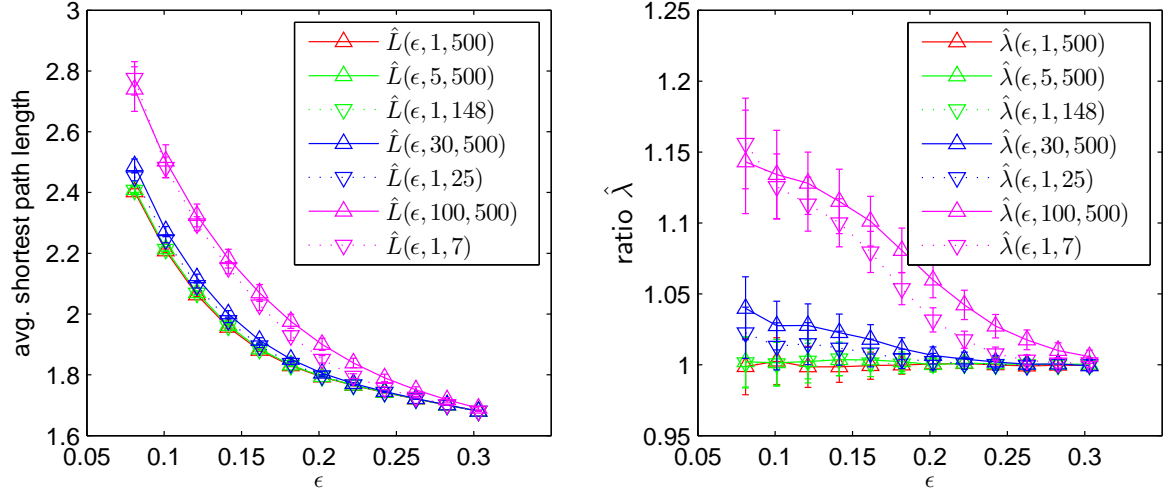
FIG. 2: **Simulation results for the average shortest path length.** Dependence of the average shortest path length $\hat{L}(\epsilon, M, T)$ (left) and of the ratio $\hat{\lambda}(\epsilon, M, T) = \hat{L}(\epsilon, 1, T)/L_{\mathrm{ER}}(\epsilon)$ (right) on edge density ϵ for different values of the size M of the moving average and of the length T of time series. Lines are for eye-guidance only.

difference becomes more pronounced the lower the edge density ϵ, the lower the length T of time series, and the larger the amount of low-frequency contributions. These results point us to an important difference between ER networks and our model networks: possible edges in ER networks are not only (1) equally likely but also (2) independently chosen to become edges. While property (1) is fulfilled in our model networks, property (2) is not.

Average shortest path length

Next we study the impact of the length of time series and of the amount of low-frequency contributions on the average shortest path length L of our model networks by employing a similar but different simulation approach as used in the previous section. To estimate L, we consider $R = 100$ networks with a fixed number of nodes ($N = 100$). We derive our model networks by thresholding $\rho_{ij,M,T}^{(r)}$, $i, j \in \{1, \ldots, N\}$, $r \in \{1, \ldots, R\}$. Let $L^{(r)}(\epsilon, 1, T)$ denote the average shortest path length for network r with $M = 1$ and different values of T, and $L^{(r)}(\epsilon, M, T')$ the average shortest path length for network r with fixed value of T ($T = T' = 500$) and different values of M. With $L_{\mathrm{ER}}^{(r)}(\epsilon)$ we refer to the average shortest path length obtained for the r-th ER network of size N and edge density ϵ. Mean values over realizations will be denoted as $\hat{L}(\epsilon, 1, T)$, $\hat{L}(\epsilon, M, T')$, and $\hat{L}_{\mathrm{ER}}(\epsilon)$ respectively. Finally, we define $\hat{\lambda}(\epsilon, 1, T) := \hat{L}(\epsilon, 1, T)/\hat{L}_{\mathrm{ER}}(\epsilon)$ and $\hat{\lambda}(\epsilon, M, T') := \hat{L}(\epsilon, M, T')/\hat{L}_{\mathrm{ER}}(\epsilon)$.

In figure 2 we show the dependence of $\hat{L}$ and $\hat{\lambda}$ on ϵ for various values of M and T. All quantities decrease as ϵ increases which can be expected due to additional edges reducing the average distances between pairs of nodes of the networks. For fixed $\epsilon \ll 1$, $\hat{L}$ takes on higher values

the higher M or the lower T. With equation (14) we have $\hat{L}(\epsilon, 1, T_{\mathrm{eff}}) \approx \hat{L}(\epsilon, M, T')$ which resembles the results obtained for the clustering coefficient. Differences between the average shortest path lengths of our model networks and ER networks (as characterized by $\hat{\lambda}$) become more pronounced the higher M and the lower T. For edge densities typically reported in field studies ($\epsilon \approx 0.1$), however, differences are less pronounced ($\hat{\lambda} \lesssim 1.2$, cf. figure 2 right) than the ones observed for the clustering coefficient ($\hat{\gamma} > 2$ for selected values of M and T, cf. figure 1 bottom left). We obtained qualitatively similar results for small ($N = 50$) and large numbers of nodes ($N = 500$).

Number of connected components and degree distribution

Since the number of connected components of a given network might affect network characteristics such as the average shortest path length (see equation (7)), we investigate the impact of different length of time series and of the amount of low-frequency contributions on the average number of connected components $\hat{N}_c(\epsilon, M, T)$ of the networks derived from $x_{i,1,T_{\mathrm{eff}}}$ and $x_{i,M,T'}$. We determine $\hat{N}_c(\epsilon, M, T)$ as the mean of R realizations of the corresponding networks and with $\hat{N}_{c,\mathrm{ER}}(\epsilon)$ we denote the mean value of the number of connected components in R realization of ER networks. For the edge densities considered here we observe ER networks to be connected (cf. figure 3), $N_{c,\mathrm{ER}} \approx 1$, which is in agreement with the connectivity condition for ER networks, $\epsilon \gg \ln N/(N-1) \approx 0.05$ (for $N = 100$). Similarly, we observe $\hat{N}_c(\epsilon, 1, T_{\mathrm{eff}}) \approx 1$, even for small values of T (cf. figure 3 right). In contrast, $\hat{N}_c(\epsilon, M, T')$ takes on higher values the lower ϵ and the higher M (cf. figure 3 left). In order to achieve a better understanding of these findings, we determine degree probability distri-
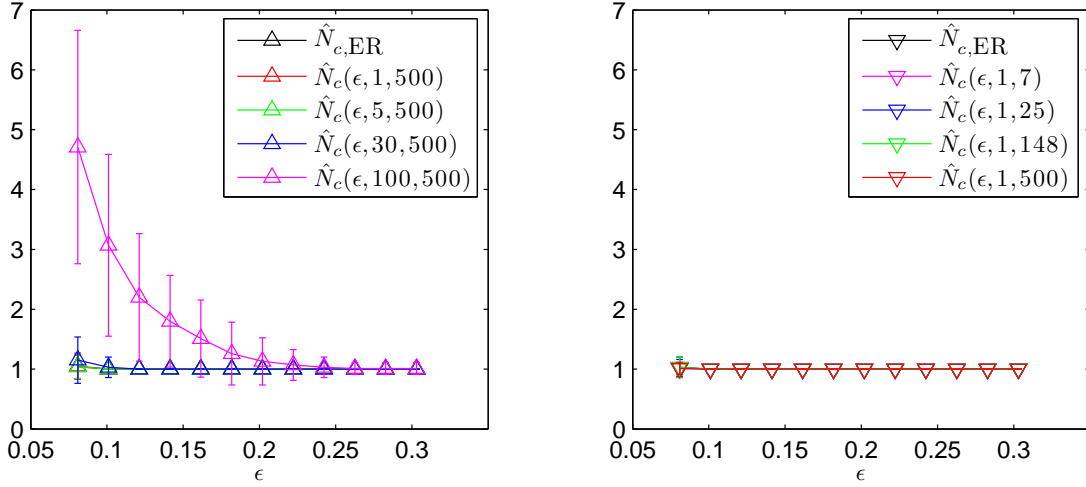
FIG. 3: **Simulation results for the number of connected components.** Dependence of the number of connected components $\hat{N}_c(\epsilon, M, T)$ on the edge density ϵ for different values of the size M of the moving average (left, for $T = 500$) and of the length T of time series (right, for $M = 1$). Lines are for eye-guidance only.

butions of our model networks. Let $\hat{p}_k$ denote the estimated probability of a node to possess a degree k, i.e., $\hat{p}_k = \#\{i^{(r)} : k_i^{(r)} = k, r \in \{1, \ldots, R\}\}/(NR)$. With $\hat{p}_k(\epsilon, M, T)$ we will denote the estimated degree distribution for networks derived from $x_{i,M,T}$. We briefly recall that the degree distribution of ER networks $p_{k,N,\text{ER}}$ follows a binomial distribution,

$$p_{k,N,\text{ER}}(\epsilon) = \binom{N-1}{k} \epsilon^k (1-\epsilon)^{N-k-1}, \qquad (15)$$

which we show in figure 4 for $N = 100$ and various ϵ (top panels and lower left panel). In the same figure we present our findings for $\hat{p}_k(\epsilon, M, T)$ for various values of $T = T_{\text{eff}}$ and M. We observe $\hat{p}_k(\epsilon, 1, T_{\text{eff}})$ to be equal to $\hat{p}_{k,N,\text{ER}}(\epsilon)$ within the error to be expected due to the limited sample size used for the estimation. For $\hat{p}_k(\epsilon, M, T')$, however, we observe striking differences in comparison to the previous degree distributions. In particular, for decreasing ϵ and higher M, the probability of nodes with degree $k = 0$ increases, which leads to networks with disconnected single nodes, thereby increasing the number of connected components of the network.

We hypothesize that the observed differences in the number of connected components as well as in the degree distributions are related to differences in the spectral content of different realizations of time series $x_{i,M,T'}^{(r)}$ for $M > 1$. In particular, a node i with a low degree k_i might be associated with a time series $x_{i,M,T'}^{(r)}$, which possesses, by chance, a small relative amount of low frequency contributions (or, equivalently, a large relative amount of high frequency contributions).

In order to test this hypothesis, we generate R realizations of $N = 100$ time series $x_{i,M,T'}^{(r)}$ and estimate their periodograms $\hat{P}_{i,M}^{(r)}(f)$ for frequencies $f \in \{0, \ldots, f_{\text{Nyq}}\}$ using a discrete Fourier transform [45]. f_{Nyq} denotes

the Nyquist frequency, and periodograms are normalized such that $\sum_f \hat{P}_{i,M}^{(r)}(f) = 1$. From the same time series, we then derive the networks using $\epsilon = 0.1$ and determine the degrees $k_i^{(r)}$. For some fixed $f' > 0$ we define the total power above f' (upper frequency range) as $\hat{P}_{i,M}^{\text{H},(r)} = \sum_{f'}^{f_{\text{Nyq}}} \hat{P}_{i,M}^{(r)}(f)$, and the total power below f' (lower frequency range) as $\hat{P}_{i,M}^{\text{L},(r)} = \sum_{f=0}^{f'-1} \hat{P}_{i,M}^{(r)}(f)$. For each realization r we estimate the correlation coefficients between the degrees and the corresponding total power contents in upper and lower frequency range, $\kappa_{\text{L}}^{(r)} = \text{corr}(k^{(r)}, \hat{P}_M^{\text{L},(r)})$ and $\kappa_{\text{H}}^{(r)} = \text{corr}(k^{(r)}, \hat{P}_M^{\text{H},(r)})$, respectively, and determine their mean values, $\kappa_{\text{L}}(M) = R^{-1} \sum_r \kappa_{\text{L}}^{(r)}$ and $\kappa_{\text{H}}(M) = R^{-1} \sum_r \kappa_{\text{H}}^{(r)}$. Note that $\kappa_{\text{L}}(M) = -\kappa_{\text{H}}(M)$ by construction. We choose $f' = f'(M)$ such that 40% of the total power of the filter function associated with the moving average is contained within the frequency range $f \in [0, f']$.

For increasing M we observe in the lower right panel of figure 4 the degrees to be increasingly correlated with $\hat{P}_M^{\text{L},(r)}$, which corresponds to an anti-correlation of degrees with $\hat{P}_M^{\text{H},(r)}$. Thus, as hypothesized above, the observed differences in the degree distributions can indeed be related to the differences in the power content of the time series. We mention that the exact choice of f' does not sensitively affect the observed qualitative relationships as long as $0 < f' \ll f_{\text{Nyq}}$ is fulfilled.

We briefly summarize the results obtained so far, which indicate a striking difference between networks derived from independent random time series using $\boldsymbol{\rho}^{\text{c}}$ or $\boldsymbol{\rho}^{\text{m}}$ (cf. equations (1) and (2)) and corresponding ER networks. First, we observed the clustering coefficient C and the average shortest path length L of our networks to be higher the lower the length T of the time series (cf. figures 1 and 2). Second, for some fixed T we observed
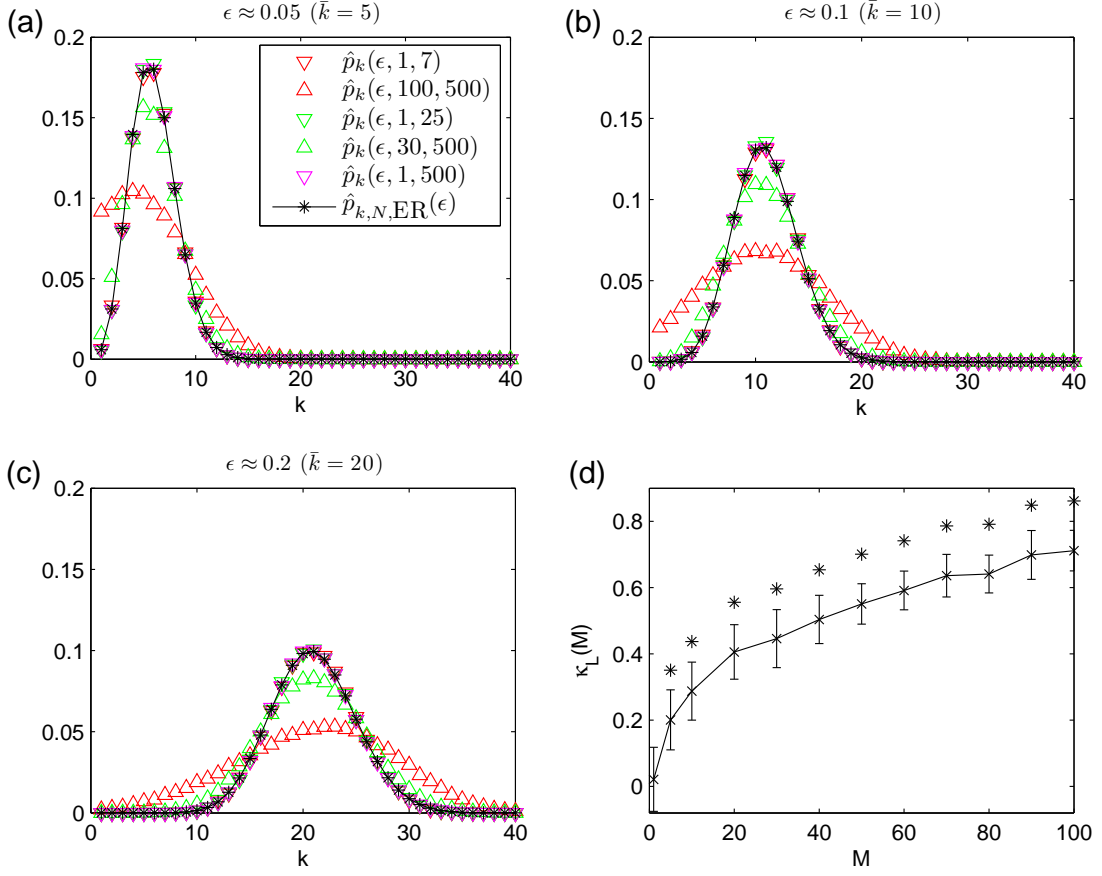
FIG. 4: **Simulation results for the degree distribution.** (**a-c**) Degree distributions $\hat{p}_k(\epsilon, M, T)$ estimated for $R = 1000$ realizations of networks derived from time series $x_{i,M,T}$ ($N = 100$) via thresholding using various edge densities $\epsilon = \bar{k}(N-1)^{-1}$ and for selected values of the size M of the moving average and of the length T of time series. The symbol legend in (a) also holds for (b) and (c). (**d**) Dependence of correlation ($\kappa_L(M)$) between node degrees and power content in the lower frequency range on the size M of the moving average. Mean values of correlations obtained for $R = 100$ realizations of networks for each value of M are shown as crosses and standard deviations as error bars. Stars indicate significant differences in comparison to $\kappa_L(1)$ (Bonferroni corrected pair-wise Wilcoxon rank sum tests for equal medians, $p < 0.01$). Lines are for eye-guidance only.

C and L to be higher the larger the amount of low frequency components (as parametrized by M) in the time series. In addition, these contributions led to an increasing number of connected components in our networks and to degree distributions that differed strongly from those of the corresponding ER networks (cf. figures 3 and 4). We mention that L as defined here (cf. equation (7)) tends to decrease for networks with an increasing number N_c of connected components, and $L \to 0$ for $N_c \to N$. L thus depends non-trivially on the amount of low frequency components in the time series. Third, for small edge densities ϵ and for short time series lengths or for a large amount of low frequency components, the clustering coefficient deviates more strongly from the one of corresponding ER networks ($\hat{\gamma} > 2$) than the average shortest path length ($\hat{\lambda} \lesssim 1.2$; cf. figure 2 right and figure 1 (bottom left)). Networks derived from independent random time series can thus be classified as small world networks if one uses $\gamma \gg 1$ and $\lambda \approx 1$ as practical criterion, which is often employed in various field studies (cf.

[31] and references therein).

Field data analysis

The findings obtained in the previous section indicate that strong low frequency contributions affect network properties C and L in a non-trivial way. We now investigate this influence in electroencephalographic (EEG) recordings of epileptic seizures that are known for their complex spatial and temporal changes in frequency content [46–49]. We analyze the multichannel ($N = 53 \pm 21$ channels) EEGs from 60 patients capturing 100 epileptic seizures reported in reference [50]. All patients had signed informed consent that their clinical data might be used and published for research purposes. The study protocol had previously been approved by the ethics committee of the University of Bonn. During the presurgical evaluation of drug-resistant epilepsy, EEG data were recorded with chronically implanted strip, grid, or
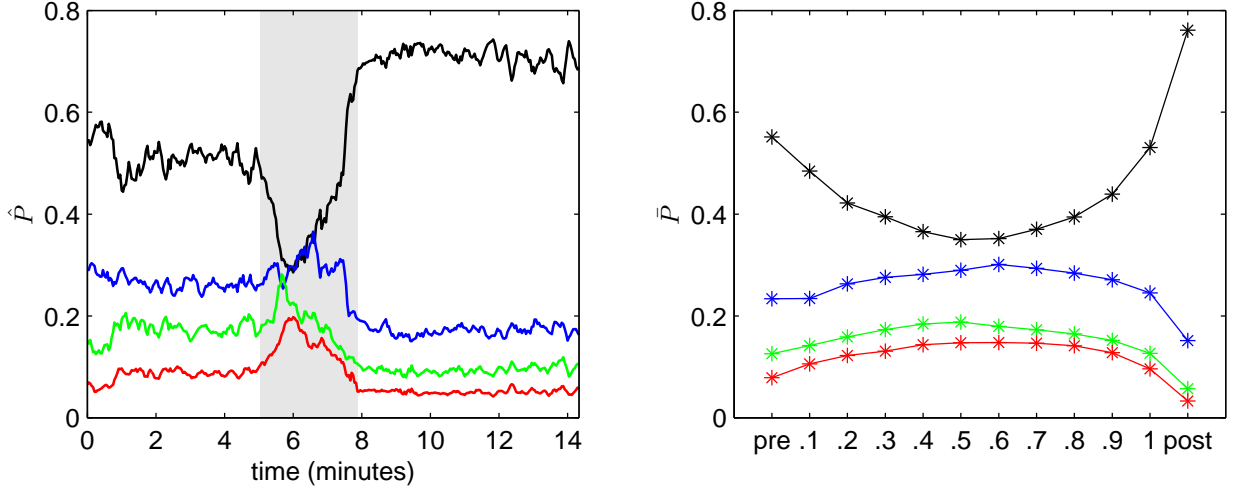
FIG. 5: **Evolving relative amount of power during epileptic seizures.** (Left) Relative amount of power $\hat{P}$ contained in the δ- ($\hat{P}_\delta$, black), ϑ- ($\hat{P}_\vartheta$, blue), α- ($\hat{P}_\alpha$, green), and β- ($\hat{P}_\beta$, red) frequency bands during an exemplary seizure. Profiles are smoothed using a four-point moving average. Grey-shaded area marks the seizure. (Right) Mean values ($\bar{P}_\delta$, $\bar{P}_\vartheta$, $\bar{P}_\alpha$, $\bar{P}_\beta$) of the relative amount of power averaged separately for pre-seizure, discretized seizure, and post-seizure time periods of 100 epileptic seizures. Lines are for eye-guidance only.

depth electrodes from the cortex and from relevant structures of the brain. The data were sampled at 200 Hz within the frequency band $0.5 - 70$ Hz using a 16-bit analog-to-digital converter. Electroencephalographic seizure onsets and seizure ends were automatically detected [51], and EEGs were split into consecutive non-overlapping windows of 2.5 s duration ($T = 500$ sampling points). Time series of each window were normalized to zero mean and unit variance. We determined $\boldsymbol{\rho}^\mathrm{c}$ and $\boldsymbol{\rho}^\mathrm{m}$ for all combinations of EEG time series from each window and derived networks with a fixed edge density $\epsilon = 0.1$ in order to exclude possible edge density effects. With L_c and C_c as well as L_m and C_m we denote characteristics of networks based on $\boldsymbol{\rho}^\mathrm{c}$ and $\boldsymbol{\rho}^\mathrm{m}$, respectively. In order to simplify matters, we omit the window indexing in the following.

We investigate a possible influence of the power content of EEG time series on the clustering coefficient and the average shortest path length by comparing their values to those obtained from ensembles of random networks that are based on properties of the EEG time series at two different levels of detail. For the first ensemble and for each patient we derive networks from random time series with a power content that approximately equals the mean power content of all EEG time series within a window. Let $\hat{P}_i(f)$ denote the estimated periodogram of each EEG time series i, and with $P(f) = N^{-1} \sum_i \hat{P}_i(f)$ we denote the mean power for each frequency component f over all N time series. We normalize $P(f)$ such that $\sum_f P(f) = 1$. We generate N random time series of length $T = 500$ whose entries are independently drawn from a uniform probability distribution, and we filter these time series in the Fourier domain using $\sqrt{P(f)}$ as filter function. We normalize the filtered time series to zero mean and unit variance and derive a network based on $\boldsymbol{\rho}^\mathrm{c}$ or $\boldsymbol{\rho}^\mathrm{m}$

($\epsilon = 0.1$). We use 20 realizations of such networks per window in order to determine the mean values of network characteristics $C_\mathrm{c}^{(1)}$ and $L_\mathrm{c}^{(1)}$ as well as $C_\mathrm{m}^{(1)}$ and $L_\mathrm{m}^{(1)}$ based on $\boldsymbol{\rho}^\mathrm{c}$ or $\boldsymbol{\rho}^\mathrm{m}$, respectively. Since the power spectra of all time series equal each other, these random networks resemble the ones investigated in the previous section.

With the second ensemble, we take into account that the power content of EEG time series recorded from different brain regions may differ substantially. For this purpose we make use of a well established method for generating univariate time series surrogates [52, 53], which have power spectral contents and amplitude distributions that are practically indistinguishable from those of EEG time series but are otherwise random. Amplitudes are iteratively permuted while the power spectrum of each EEG time series is approximately preserved. Since this randomization scheme destroys any significant linear or non-linear dependencies between time series, it has been successfully applied to test the null hypothesis of independent linear stochastic processes. For each patient, we generated 20 surrogate time series for each EEG time series from each recording site and each window, and derived networks based on either $\boldsymbol{\rho}^\mathrm{c}$ or $\boldsymbol{\rho}^\mathrm{m}$ ($\epsilon = 0.1$). Mean values of characteristics of these random networks are denoted as $C_\mathrm{c}^{(2)}$ and $L_\mathrm{c}^{(2)}$ as well as $C_\mathrm{m}^{(2)}$ and $L_\mathrm{m}^{(2)}$, respectively.

We begin with an exemplary recording of a seizure of which we show in figure 5 (left) the temporal evolution of the relative amount of power in the δ- (0–4 Hz, P_δ), ϑ- (4–8 Hz, P_ϑ), α- (8–12 Hz, P_α), and β- (12–20 Hz, P_β) frequency bands. Prior to the seizure the δ-band contains more than 50% of the total power which is then shifted towards higher frequencies and back towards low frequencies at seizure end. P_δ is even higher after the seizure than prior to the seizure.
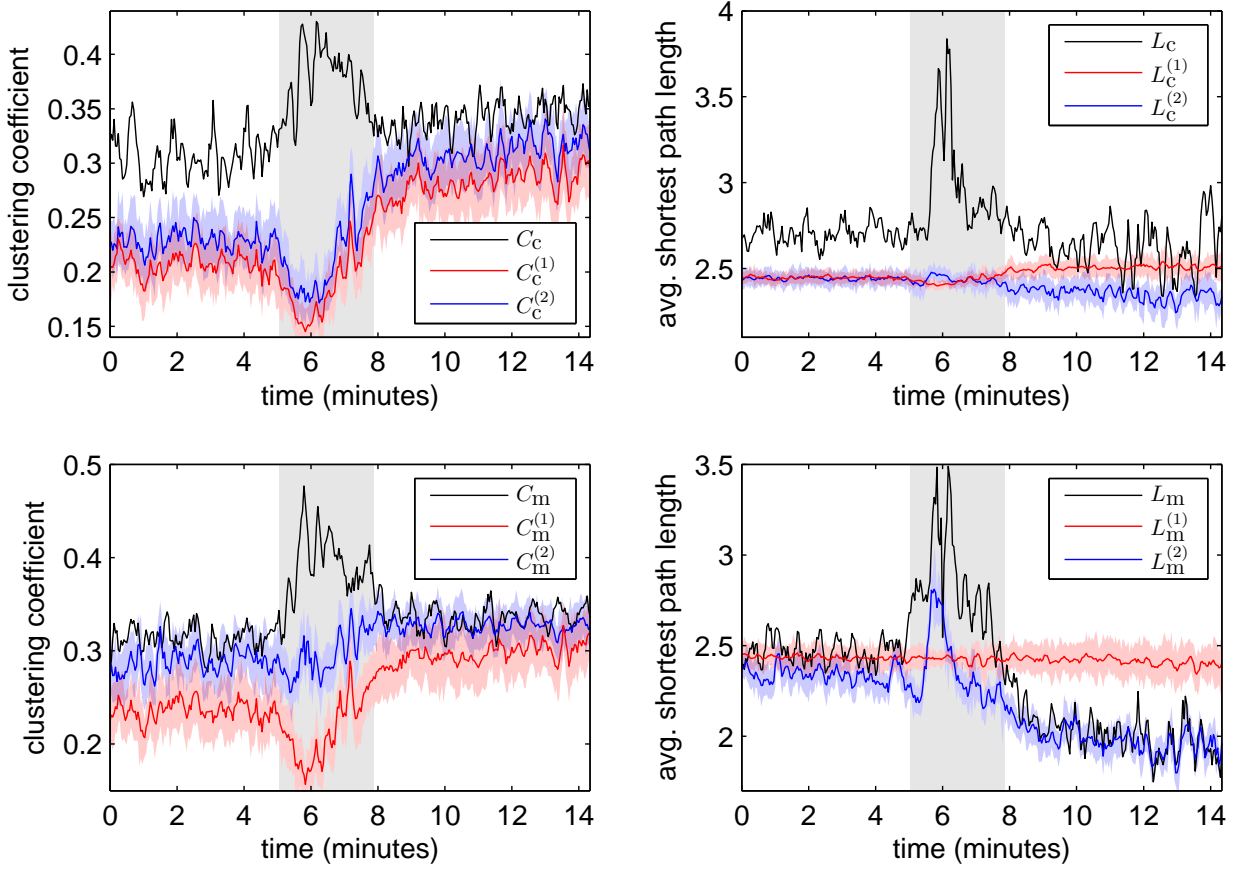
FIG. 6: **Evolving network properties during an exemplary epileptic seizure.** Network properties C_c and L_c (top row, black lines) as well as C_m and L_m (bottom row, black lines) during an exemplary seizure (cf. figure 5 (left)). Mean values and standard deviations of network properties obtained from surrogate time series ($C_c^{(2)}$, $L_c^{(2)}$, $C_m^{(2)}$, $L_m^{(2)}$) are shown as blue lines and blue shaded areas, respectively, and mean values and standard deviations of network properties obtained from the overall power content model ($C_c^{(1)}$, $L_c^{(1)}$, $C_m^{(1)}$, $L_m^{(1)}$) are shown as red lines and red shaded areas, respectively. Profiles are smoothed using a four-point moving average. Grey-shaded area marks the seizure. For corresponding Erdős-Rényi networks $C_{ER} \approx 0.1$ and $L_{ER} \approx 2.4$ for all time windows.

In figure 6 we show the temporal evolution of network properties obtained for this recording based on $\boldsymbol{\rho}^c$ (top panels) and $\boldsymbol{\rho}^m$ (bottom panels). During the seizure both the clustering coefficients C_c and C_m and the average shortest path lengths L_c and L_m show pronounced differences to the respective properties obtained from the random networks. These differences are less pronounced prior to and after the seizure, where $C_m^{(2)}$ and $L_m^{(2)}$ even approach the values of C_m and L_m, respectively. $C_c^{(1)}$ and $C_m^{(1)}$ decrease during the seizure and already increase prior to seizure end, resembling the changes of P_δ (cf. figure 5 (left)). This is in accordance with results of our simulation studies: there we observed the clustering coefficient to be higher the larger the amount of low frequency components in the time series; this could also be observed, but to a much lesser extent, for the average shortest path length. Indeed, $L_m^{(1)}$ and $L_c^{(1)}$ vary little over time, and $L_c^{(1)}$ is only slightly increased after the seizure, reflecting the high amount of power in the

δ-band.

We only observe small deviations between $C_c^{(1)}$ and $C_c^{(2)}$ as well as between $L_c^{(1)}$ and $L_c^{(2)}$, which appear to be systematic (for many windows $C_c^{(1)} \lesssim C_c^{(2)}$ and $L_c^{(1)} \gtrsim L_c^{(2)}$). These suggest that for interaction networks derived from $\boldsymbol{\rho}^c$, both random network ensembles appear appropriate to characterize the influence of power in low frequency bands on clustering coefficient and the average shortest path length. In contrast, we observed differences between $C_m^{(1)}$ and $C_m^{(2)}$, as well as between $L_m^{(1)}$ and $L_m^{(2)}$. These differences were most pronounced during the seizure and for $L_m^{(1)}$ and $L_m^{(2)}$ also after the seizure. This finding indicates that the clustering coefficient and average shortest path length of interaction networks derived from $\boldsymbol{\rho}^m$ depend sensitively on the power contents of EEG time series recorded from different brain regions. Thus, for these interaction networks only the random networks that account for the complex changes in frequency content of different brain regions prior to, during, and after
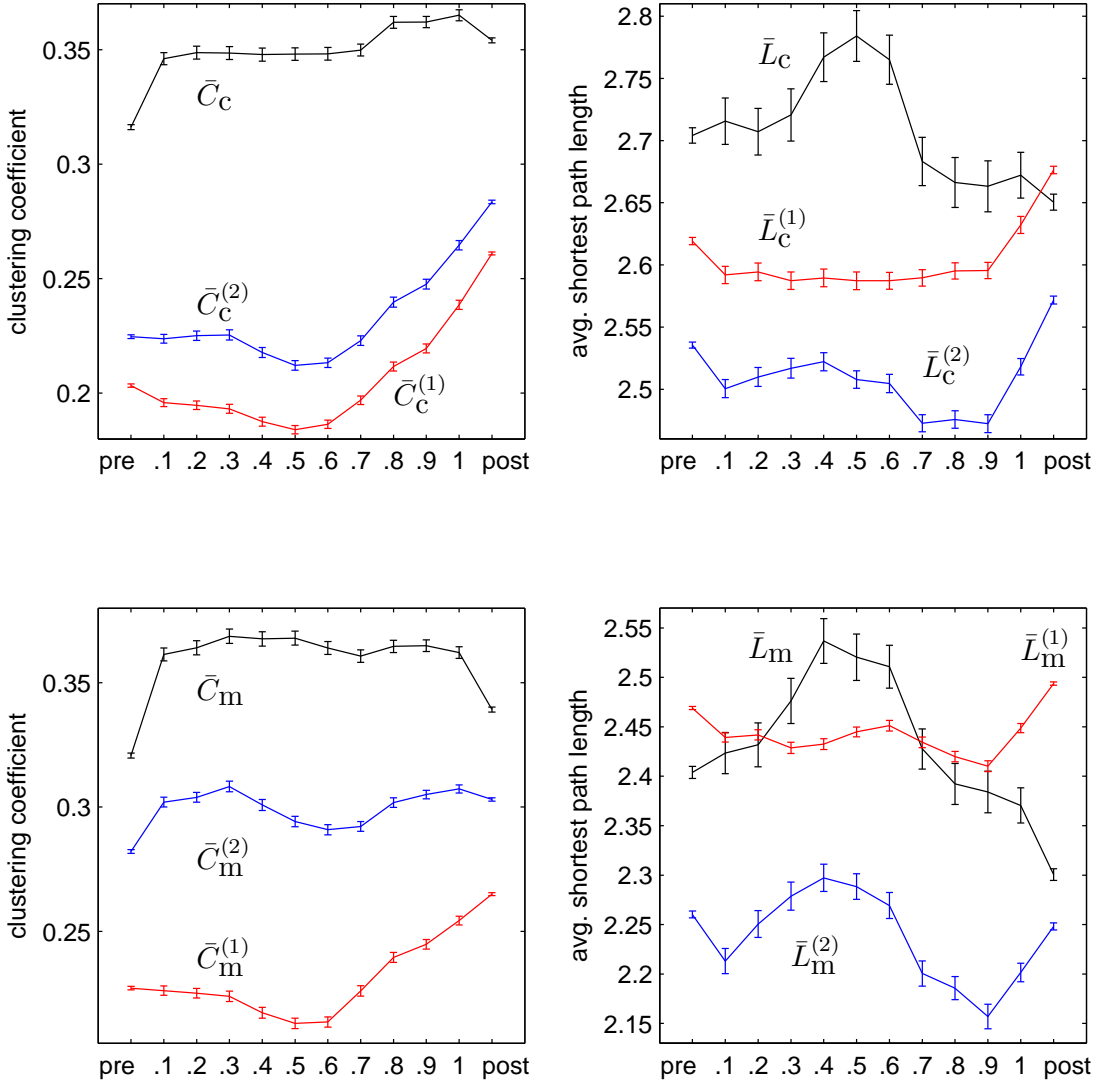
FIG. 7: **Evolving network properties averaged over 100 epileptic seizures.** Mean values (black) of network properties C_c (top left), L_c (top right), C_m (bottom left), and L_m (bottom right) averaged separately for pre-seizure, discretized seizure, and post-seizure time periods of 100 epileptic seizures. Mean values of corresponding network properties obtained from the first and the second ensemble of random networks are shown as red and blue lines, respectively. All error bars indicate standard error of the mean. Lines are for eye-guidance only.

seizures appear appropriate to characterize the influence of power in low frequency bands on clustering coefficient and the average shortest path length.

We continue by studying properties of networks derived from the EEG recordings of all 100 focal onset seizures. Due to the different durations of seizures (mean seizure duration: 110 ± 60 s) we partitioned each seizure into 10 equidistant time bins (see reference [50] for details) and assigned the time-dependent network properties to the respective time bins. For each seizure we included the same number of pre-seizure and post-seizure windows in our analysis and assigned the corresponding time-dependent network properties to one pre-seizure and one post-seizure time bin. Within each time bin we determined the mean value (e.g., $\bar{C}_c$) and its standard

error for each property. In figure 5 (right), we show for each time bin the mean values of the relative amount of power in different frequency bands of all seizure recordings ($\bar{P}_\delta$, $\bar{P}_\vartheta$, $\bar{P}_\alpha$, $\bar{P}_\beta$). Similar to the exemplary recording (cf. figure 5 (left)), we observe a shift in the relative amount of power in low frequencies prior to seizures towards higher frequencies during seizures and back to low frequencies at seizure end. The amount of power in the δ-band is on average higher in the post-seizure bin than in the pre-seizure bin.

In figure 7 we show the mean values of properties of networks in each time bin for all seizures. We observe $\bar{C}_c^{(1)}$, $\bar{C}_c^{(2)}$, $\bar{L}_c^{(1)}$, $\bar{L}_c^{(2)}$, $\bar{C}_m^{(1)}$, and $\bar{L}_m^{(1)}$ to decrease during seizures and to increase prior to seizure end thereby roughly reflecting the amount of power contained in low
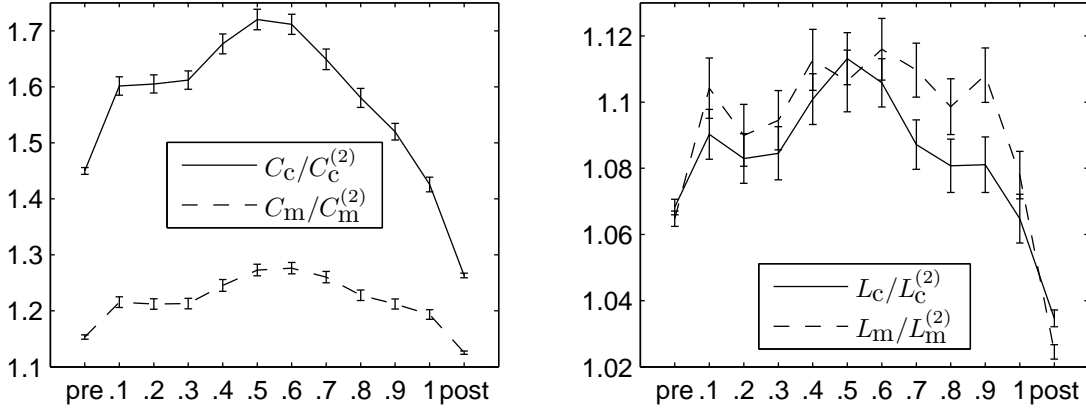
FIG. 8: **Evolving normalized network properties averaged over 100 epileptic seizures.** Mean values of $C_c/C_c^{(2)}$ and $C_m/C_m^{(2)}$ (left) as well as $L_c/L_c^{(2)}$ and $L_m/L_m^{(2)}$ (right) averaged separately for pre-seizure, discretized seizure, and post-seizure time periods of 100 epileptic seizures. All error bars indicate standard error of the mean. Lines are for eye-guidance only.

frequencies (cf. figure 5 (right), $\bar{P}_\delta$). $\bar{C}_c^{(1)}$ and $\bar{C}_c^{(2)}$ and to a lesser extent also $\bar{L}_c^{(1)}$ and $\bar{L}_c^{(2)}$ roughly follow the same course in time, however, with a slight shift in the range of values as already observed in the exemplary recording of a seizure (cf. figure 6). Differences between both random network ensembles are most pronounced in network properties based on $\boldsymbol{\rho}^m$, i.e., between $\bar{C}_m^{(1)}$ and $\bar{C}_m^{(2)}$ as well as between $\bar{L}_m^{(1)}$ and $\bar{L}_m^{(2)}$. This corroborates the observation that the clustering coefficient and the average shortest path length of the random networks based on $\boldsymbol{\rho}^m$ depend more sensitively on the power contents of EEG time series recorded from different brain regions than the respective quantities derived from $\boldsymbol{\rho}^c$. While $\bar{L}_c$ and $\bar{L}_m$ show a similar course in time, reaching a maximum in the middle of the seizures, we observe a remarkable difference between $\bar{C}_c$ and $\bar{C}_m$ prior to end of the seizures, where the amount of power in low frequencies is large. While $\bar{C}_m$ decreases at the end of the seizures, $\bar{C}_c$ does not and remains elevated after seizures. Interestingly, considering the corresponding quantities obtained from the second random network ensemble, $\bar{C}_m^{(2)}$ fluctuates around 0.3 ± 0.01 and does not increase at the end of seizures, while, in contrast, $\bar{C}_c^{(2)}$ increases at the end of the seizures, traversing an interval of values roughly three times larger than the interval containing values of $\bar{C}_m^{(2)}$. Taken together these findings suggest that the pronounced changes of the frequency content of EEG time series seen during epileptic seizures influence the values of the clustering coefficient and the average shortest path length.

A comparison of some value of a network property with the one obtained for a random network with the same edge density and number of nodes is typically achieved by calculating their ratio. If ER networks are used for comparison, the value of a network property is rescaled by a constant factor. In this case, the time-dependent changes of network properties shown in figure 7 will be shifted along the ordinate only. In order to take into

account the varying power content of EEG time series recorded from different brain regions we instead normalize the clustering coefficients and the average shortest path lengths with the corresponding quantities from the second random network ensemble $\bar{C}_c^{(2)}$, $\bar{C}_m^{(2)}$, $\bar{L}_c^{(2)}$, and $\bar{L}_m^{(2)}$ (cf. figure 8). We observe the normalized network properties to describe a concave-like movement over time indicating a reconfiguration of networks from more random (before seizures) towards a more regular (during seizures) and back towards more random network topologies. This is in agreement with previous observations using a different and seldom used thresholding method [50].

Discussion

The network approach towards the analysis of empirical multivariate time series is based on the assumption that the data is well represented by a model of mutual relationships (i.e., a network). We studied interaction networks derived from finite time series generated by independent processes that would not advocate a representation by a model of mutual relationships. We observed the derived interaction networks to show non-trivial network topologies. These are induced by the finiteness of data, which limits reliability of estimators of signal interdependence, together with the use of a frequently employed thresholding technique. Since the analysis methodology alone can already introduce non-trivial structure in the derived networks, the question arises as to how informative network analysis results obtained from finite empirical data are with respect to the studied dynamics. This question may be addressed by defining and making use of appropriate null models. In the following, we briefly discuss two null models that are frequently employed in field studies.

Erdős-Rényi (ER) networks represent one of the earliest and best studied network models in mathematical

literature and can be easily generated. They can be used to test whether the network under consideration complies with the notion of a random network in which possible edges are equally likely and independently chosen to become edges. We observed that clustering coefficient C and average shortest path length L for interaction networks derived from finite random time series differed pronouncedly from those obtained from corresponding ER networks, which would likely lead to a classification of interaction networks as small-world networks. Since the influence of the analysis methodology is not taken into account with ER networks, they may not be well suited for serving as null models in studies of interaction networks derived from finite time series.

Another null model is based on randomizing the network topology while preserving the degrees of nodes [26, 54, 55]. It is used to evaluate whether the network under consideration is random under the constraint of a given degree sequence. Results of our simulation studies point out that the structures induced in the network topology by the way how networks are derived from empirical time series cannot be related to the degree sequence only. We observed that C and L from interaction networks remarkably depended on the finiteness of the data, while the degree distribution did not (cf. figure 4 (a-c), $M = 1$). The usefulness of degree-preserving randomized networks has also been subject of debate since they do not take into account different characteristics of the data and its acquisition [56, 57]. Moreover, the link-switching algorithm frequently employed for generating such networks has been shown to non-uniformly sample the space of networks with predefined degree sequence (see, e.g., references [25, 58]). This deficiency can be addressed by using alternative randomization schemes (see, e.g., [58–60] and references therein).

We propose to take into account the finite length and the frequency contents of time series when defining null models. For this purpose we applied the same methodological steps as in field data analysis (estimation of signal interdependence and thresholding of interdependence values to define links) but used surrogate time series [53] to derive random networks (second ensemble). These surrogate time series comply with the null hypothesis of independent linear stochastic processes and preserve length, frequency content, and amplitude distribution of the original time series. For these random networks, we observed (in our simulation studies) dependencies between properties of networks and properties of time series: the clustering coefficient C, and, to a lesser extent, the average shortest path length L are higher the higher the relative amount of low frequency components, the shorter the length of time series, and the smaller the edge density of the network. Results obtained from an analysis of interaction networks derived from multichannel EEG recordings of one hundred epileptic seizures confirm that the pronounced changes of the frequency content seen during seizures influence the values of C and L. Comparing these network characteristics with those obtained from our random networks allowed us to distinguish aspects of global network dynamics during seizures from those spuriously induced by the applied methods of analysis.

Our random networks will likely be classified as small-world networks when compared to ER networks which might indicate that small-world topologies in networks derived from empirical data as reported in an ever increasing number of studies can partly or solely be related to the finite length and frequency content of time series. If so, small-world topologies would be an overly complicated description of the simple finding of finite time series with a large amount of low frequency components. In this context, our approach could be of particular interest for studies that deal with short time series and low frequency contents, as, for example, is the case in resting state functional magnetic resonance imaging studies (see, e.g., references [61–65]). In such studies, taking into account potential frequency effects could help to unravel information on the network level that would be otherwise masked.

We observed the degrees of nodes of our random networks to be correlated with the relative amount of power in low-frequencies in the respective time series (cf. figure 4). The degree of a node has been used in field studies as an indicator of its centrality in the network (see, e.g., [2, 66] and references therein). Particular interest has been devoted to nodes which are highly central (hubs). In this context it would be interesting to study whether findings of hubs in interaction networks can partly or solely be explained by the various frequency contents of time series entering the analysis. In such a case, hubs would be a complicated representation of features already present on a single time series level. We are confident that our random networks can help to clarify this issue.

Our simulation studies were based on the simplified assumption that power spectra of all time series from which a network is derived are approximately equal. The dependencies of C and L on the power content could also be observed qualitatively for networks derived from EEG time series – that were recorded from different brain regions and whose power spectra may differ substantially among each other – but only if link definition was based on thresholding the values of the correlation coefficient (ρ^{c}). Thus, estimating mean power spectra of multivariate time series can provide the experimentalist with a rule of thumb for the potential relative increase of C and L in different networks based on the correlation coefficient. This rule of thumb, however, might not be helpful if the maximum value of the absolute cross correlation (ρ^{m}) is used to estimate signal interdependencies. In this case, C and L depended sensitively on the heterogeneity of power spectra (see the second random network ensemble). It would be interesting to investigate in future studies, which particular properties of ρ^{c} and ρ^{m} can be accounted for these differences.

We close the discussion with two remarks, the first being of interest for experimentalists. Our findings also

shed light on a network construction technique that relies on significance testing in order to decide upon defining a link or not [21]. For this purpose, a null distribution of a chosen estimator of signal interdependence (ρ^{m}) is generated for each pair of time series and a link is established if the null hypothesis of independent processes generating the time series can be rejected at a predefined significance level. It was suggested in Ref. [21] to use a limited subset of time series in order to minimize computational burden when generating null distributions. Our findings indicate that networks constructed this way will yield an artificially increased number of false positive or of false negative links which will depend on the frequency contents of time series being part or not part of the subset. Our second remark is related to network modeling. By choosing some threshold and generating time series that satisfy the relation between the size of the moving average and the length of time series, networks can be generated which differ in their degree distributions but approximately equal in their clustering coefficient and average shortest path length. This property could be of value for future modeling studies.

To summarize, we have demonstrated that interaction networks, derived from finite time series via thresholding an estimate of signal interdependence, can exhibit non-trivial properties that solely reflect the mostly unavoidable finiteness of empirical data, which limits the reliability of signal interdependence estimators. Addressing these influences, we proposed random network models that take into account the way interaction networks are derived from the data. With an exemplary time-resolved analysis of the clustering coefficient C and the average shortest path length L of interaction networks derived from multichannel electroencephalographic recordings of one hundred epileptic seizures, we demonstrated that our random networks allow one to gain deeper insights into the global network dynamics during seizures. Here we concentrated on C and L but we also expect other network characteristics to be influenced by the methodologies used to derive interaction networks from empirical data. Analytical investigations of properties of our random networks and the development of formal tests for deviations from these networks may be regarded as promising topics for further studies. Other research directions are related to the framework we proposed to generate random networks from time series. For example, parts of the framework may be exchanged in order to study network construction methodologies other than thresholding (e.g., based on minimum spanning trees [14] or based on allowing weighted links) or other widely used linear and nonlinear methods for estimating signal interdependence [34, 35, 39]. Other surrogate concepts [67–72] may allow for defining different random networks tailored to various purposes. We believe that research into network inference from time series and into random network models that incorporate knowledge about the way networks are derived from empirical data can decisively advance applied network science. This line of research can contribute to gain a better understanding of complex dynamical systems studied in various scientific fields.

Appendix A

Lemma 1

For every $i, j \in \{1, \ldots, N\}$ with $i \neq j$, we have the following limit of the probability distribution of the empirical correlation:

$$P\left(\sqrt{\frac{T}{g(M)}}\mathrm{corr}(x_{i,M,T}, x_{j,M,T}) \leq x\right) \to \Phi(x) \quad \text{(A1)}$$

with

$$g(M) = \frac{2}{3}M + \frac{1}{3}\frac{1}{M}$$

as $T \to \infty$, where Φ denotes the cumulative distribution function of a standard normal random variable.
Proof. In order to simplify the presentation, we write $y_{i,M,T}(t) = x_{i,M,T}(t) - \frac{1}{2}$, so that $Ey_{i,M,T}(t) = 0$. First note that $y_{i,M,T}(t)$ is a M-dependent sequence, i.e. for $|s - t| > M$, $y_{i,M,T}(s)$ and $y_{i,M,T}(t)$ are independent. So we have that the covariance

$$\mathrm{Cov}\left(y_{i,M,T}(1)y_{j,M,T}(1), y_{i,M,T}(t)y_{j,M,T}(t)\right) = 0$$

for $T > M$. Additionally,

$$\mathrm{Cov}\left(y_{i,M,T}(1)y_{j,M,T}(1), y_{i,M,T}(t)y_{j,M,T}(t)\right) =$$
$$\mathrm{Cov}\left(y_{i,M,T}(1), y_{i,M,T}(t)\right)\mathrm{Cov}\left(y_{j,M,T}(1), y_{j,M,T}(t)\right) \quad \text{(A2)}$$

and $\mathrm{Cov}\left(z_i(s), z_i(t)\right) = \mathrm{Var}\left(z_i(1)\right)$ if $s = t$ and otherwise $\mathrm{Cov}\left(z_i(s), z_i(t)\right) = 0$. For $1 \leq t \leq M$, we obtain by the definition of the moving average and the independence of the underlying process $z_j(t)$, $t \in \mathbb{N}$ that

$$\mathrm{Cov}\left(y_{i,M,T}(1)y_{j,M,T}(1), y_{i,M,T}(t)y_{j,M,T}(t)\right) \quad \text{(A3)}$$

$$= \frac{1}{M^4}\left(\sum_{s=1}^{M-(t-1)}\mathrm{Var}\left(z_j(s)\right)\right)^2 \quad \text{(A4)}$$

$$= \frac{1}{M^4}(M - (t-1))^2\mathrm{Var}^2\left(z_i(1)\right). \quad \text{(A5)}$$

By the central limit theorem for M-dependent random variables, see reference [73],

$$\frac{1}{\sqrt{\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}y_{i,M,T}(t)y_{j,M,T}(t)\right)}}$$
$$\times \frac{1}{T}\sum_{t=1}^{T}y_{i,M,T}(t)y_{j,M,T}(t) \quad \text{(A6)}$$

converges in distribution to a standard normal random varibale as $T \to \infty$. Furthermore, we have the following convergence for the variance as $T \to \infty$:

$$TVar\left(\frac{1}{T}\sum_{t=1}^{T}y_{i,M,T}(t)y_{j,M,T}(t)\right)$$

$$\rightarrow Var(y_{i,M,T}(1)y_{j,M,T}(1)) + 2\sum_{t=2}^{M}Cov\left(y_{i,M,T}(1)y_{j,M,T}(1), y_{i,M,T}(t)y_{j,M,T}(t)\right)$$

$$=\left(\frac{1}{M^2}+\frac{2}{M^4}\sum_{t=2}^{M}(M-(t-1))^2\right)Var^2\left(z_i(1)\right) = \frac{g(M)}{M^2}Var^2\left(z_i(1)\right). \quad (A7)$$

The last equality follows easily by $\sum_{i=1}^{n}i^2 = \frac{n(n+1)(2n+1)}{6}$. With the same central limit theorem, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_{i,M,T}(t)$ converges to a normal limit, so $\frac{1}{T^{\frac{3}{4}}}\sum_{t=1}^{T}y_{i,M,T}(t) \rightarrow 0$ in probability and consequently

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}y_{i,M,T}(t)\right)\left(\frac{1}{T}\sum_{t=1}^{T}y_{j,M,T}(t)\right)$$

$$=\left(\frac{1}{T^{\frac{3}{4}}}\sum_{t=1}^{T}y_{i,M,T}(t)\right)\left(\frac{1}{T^{\frac{3}{4}}}\sum_{t=1}^{T}y_{j,M,T}(t)\right) \rightarrow 0 \quad (A8)$$

in probability as $T \rightarrow \infty$. By similar arguments, we have that $\frac{1}{T}\sum_{t=1}^{T}y_{i,M,T}^2(t) \rightarrow Var(y_{i,M,T}(1)) = \frac{1}{M}Var\left(z_i(1)\right)$

and $\frac{1}{T}\sum_{t=1}^{T}y_{i,M,T}(t) \rightarrow 0$, so we get

$$\frac{1}{T}\sum_{t=1}^{T}(y_{i,M,T}(t)-\bar{y}_{i,M,T})^2 =$$

$$\frac{1}{T}\sum_{t=1}^{T}y_{i,M,T}^2(t) - \left(\frac{1}{T}\sum_{t=1}^{T}y_{i,M,T}(t)\right)^2$$

$$\rightarrow Var(y_{i,M,T}(1)) = \frac{1}{M}Var\left(z_i(1)\right). \quad (A9)$$

By Slutsky's theorem [74] and with (A6), (A7), (A8), and (A9), we finally obtain that

$$\sqrt{\frac{T}{g(M)}}corr(x_{i,M,T}, x_{j,M,T}) = \frac{\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}y_{i,M,T}(t)y_{j,M,T}(t) - \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}y_{i,M,T}(t)\right)\left(\frac{1}{T}\sum_{t=1}^{T}y_{j,M,T}(t)\right)}{\sqrt{g(M)\frac{1}{T}\sum_{t=1}^{T}(y_{i,M,T}(t)-\bar{y}_{i,M,T})^2\frac{1}{T}\sum_{t=1}^{T}(y_{j,M,T}(t)-\bar{y}_{j,M,T})^2}} \quad (A10)$$

converges in distribution to a standard normal random variable as $T \rightarrow \infty$. This completes the proof.

in probability with $T_{\text{eff}}(M) = \frac{T}{g(M)}$.

Lemma 2

For $T \rightarrow \infty$, $R \rightarrow \infty$

$$\hat{\epsilon}\left(\frac{\theta}{\sqrt{T_{\text{eff}}(M)}}, M, T\right) \rightarrow 2\Phi(-\theta) \quad (A11)$$

Proof. With Lemma 1, we have that

$$E\left[H^{(r)}_{ij,M,T}\left(\frac{\theta}{\sqrt{T_{\text{eff}}(M)}}\right)\right] = P\left(\rho_{ij,M,T} > \frac{\theta}{\sqrt{T_{\text{eff}}(M)}}\right)$$

$$= P\left(\text{corr}(x_{i,M,T}, x_{j,M,T}) > \frac{\theta}{\sqrt{T_{\text{eff}}(M)}}\right) + P\left(\text{corr}(x_{i,M,T}, x_{j,M,T}) < \frac{-\theta}{\sqrt{T_{\text{eff}}(M)}}\right)$$

$$= P\left(\sqrt{\frac{T}{g(M)}}\rho_{ij,M,T} > \theta\right) + P\left(\sqrt{\frac{T}{g(M)}}\rho_{ij,M,T} < -\theta\right) \to 2\Phi(-\theta)$$

as $T \to \infty$. Furthermore, $H^{(r)}_{ij,M,T}$ is bounded by 0 and 1, so $\text{Var}\left(H^{(r)}_{ij,M,T}\right) \le \frac{1}{4}$. By the independence of the R random networks

$$\text{Var}\left(\hat{\epsilon}\left(\frac{\theta}{\sqrt{T_{\text{eff}}(M)}}, M, T\right)\right)$$

$$= \frac{1}{R^2}\sum_{r=1}^{R}\text{Var}\left(H^{(r)}_{ij,M,T}\left(\frac{\theta}{\sqrt{T_{\text{eff}}(M)}}\right)\right) \le \frac{1}{4R} \to 0 \tag{A12}$$

as $R \to \infty$. The lemma follows with the Chebyshev in-equality.

[1] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[3] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, Phys. Rep. **469**, 93 (2008).
[4] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, New York, USA, 2008).
[5] A. A. Tsonis and P. J. Roebber, Physica A **333**, 497 (2004).
[6] K. Yamasaki, A. Gozolchiani, and S. Havlin, Phys. Rev. Lett. **100**, 228501 (2008).
[7] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, Eur. Phys. J.–Spec. Top. **174**, 157 (2009).
[8] A. A. Tsonis, G. Wang, K. L. Swanson, F. A. Rodrigues, and L. da Fontura Costa, Clim. Dynam. **37**, 933 (2011).
[9] K. Steinhaeuser, N. V. Chawla, and A. R. Ganguly, Statistical Analysis and Data Mining **4**, 497 (2011).
[10] S. Abe and N. Suzuki, Physica A **337**, 357 (2004).
[11] S. Abe and N. Suzuki, Nonlinear Proc. Geoph. **13**, 145 (2006).
[12] A. Jiménez, K. F. Tiampo, and A. M. Posadas, Nonlinear Proc. Geoph. **15**, 389 (2008).
[13] T. R. Krishna Mohan and P. G. Revathi, J. Seismol. **15**, 71 (2011).
[14] R. N. Mantegna, Eur. Phys. J. B **11**, 193 (1999).
[15] J. P. Onnela, K. Kaski, and J. Kertesz, Eur. Phys. J. B **38**, 353 (2004).
[16] V. Boginski, S. Butenko, and P. M. Pardalos, Comput. Stat. An. **48**, 431 (2005).
[17] T. Qiu, B. Zheng, and G. Chen, New J. Physics **12**, 043057 (2010).
[18] F. Emmert-Streib and M. Dehmer, PLoS ONE **5**, e12884 (2010).
[19] J. C. Reijneveld, S. C. Ponten, H. W. Berendse, and C. J. Stam, Clin. Neurophysiol. **118**, 2317 (2007).
[20] E. Bullmore and O. Sporns, Nat. Rev. Neurosci. **10**, 186 (2009).
[21] M. A. Kramer, U. T. Eden, S. S. Cash, and E. D. Kolaczyk, Phys. Rev. E **79**, 061916 (2009).
[22] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, Europhys. Lett. **87**, 48007 (2009).
[23] F. Emmert-Streib and M. Dehmer, Complexity **16**, 24 (2010).
[24] P. Erdős and A. Rényi, Publ. Math. Debrecen **6**, 290 (1959).
[25] A. R. Rao, R. Jana, and S. Bandyopadhyay, Sankhya Ser. A **58**, 225 (1996).
[26] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[27] R. James, D. P. Croft, and J. Krause, Behav. Ecol. Sociobiol. **63**, 989 (2009).
[28] G. Lima-Mendez and J. van Helden, Mol. Biosyst. **5**, 1482 (2009).
[29] A. A. Ioannides, Curr. Opin. Neurobiol. **17**, 161 (2007).
[30] C. T. Butts, Science **325**, 414 (2009).
[31] S. Bialonski, M.-T. Horstmann, and K. Lehnertz, Chaos **20**, 013134 (2010).
[32] L. Antiqueira, F. A. Rodrigues, B. C. M. van Wijk, L. da F. Costa, and A. Daffertshofer, NeuroImage **53**, 439 (2010).

[33] F. Gerhard, G. Pipa, B. Lima, S. Neuenschwander, and W. Gerstner, Front. Comp. Neuroscience **5**, 4 (2011).

[34] D. Brillinger, *Time Series: Data Analysis and Theory* (Holden-Day, San Francisco, USA, 1981).

[35] A. S. Pikovsky, M. G. Rosenblum, and J. Kurths, *Synchronization: A universal concept in nonlinear sciences* (Cambridge University Press, Cambridge, UK, 2001).

[36] S. Boccaletti, J. Kurths, G. Osipov, D. L. Valladares, and C. S. Zhou, Phys. Rep. **366**, 1 (2002).

[37] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, UK, 2003), 2nd ed.

[38] E. Pereda, R. Quian Quiroga, and J. Bhattacharya, Prog. Neurobiol. **77**, 1 (2005).

[39] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007).

[40] K. Lehnertz, S. Bialonski, M.-T. Horstmann, D. Krug, A. Rothkegel, M. Staniek, and T. Wagner, J. Neurosci. Methods **183**, 42 (2009).

[41] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[42] V. Latora and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001).

[43] V. Latora and M. Marchiori, Eur. Phys. J. B **32**, 249 (2003).

[44] F. Chung and L. Lu, Adv. Appl. Math. **26**, 257 (2001).

[45] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, Cambridge, UK, 2002), 2nd ed.

[46] P. J. Franaszczuk, G. K. Bergey, P. J. Durka, and H. M. Eisenberg, Electroencephalogr. Clin. Neurophysiol. **106**, 513 (1998).

[47] S. J. Schiff, D. Colella, G. M. Jacyna, E. Hughes, J. W. Creekmore, A. Marshall, M. Bozek-Kuzmicki, G. Benke, W. D. Gaillard, J. Conry, et al., Clin. Neurophysiol. **111**, 953 (2000).

[48] C. C. Jouny, P. J. Franaszczuk, and G. K. Bergey, Clin. Neurophysiol. **114**, 426 (2003).

[49] F. Bartolomei, D. Cosandier-Rimele, A. McGonigal, S. Aubert, J. Regis, M. Gavaret, F. Wendling, and P. Chauvel, Epilepsia **51**, 2147 (2010).

[50] K. Schindler, S. Bialonski, M.-T. Horstmann, C. E. Elger, and K. Lehnertz, Chaos **18**, 033119 (2008).

[51] K. Schindler, H. Leung, C. E. Elger, and K. Lehnertz, Brain **130**, 65 (2007).

[52] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996).

[53] T. Schreiber and A. Schmitz, Physica D **142**, 346 (2000).

[54] J. M. Roberts, Soc. Networks **22**, 273 (2000).

[55] S. Maslov, K. Sneppen, and A. Zaliznyak, Physica A **333**, 529 (2004).

[56] Y. Artzy-Randrup, S. J. Fleishman, N. Ben-Tal, and L. Stone, Science **305**, 1107 (2004).

[57] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, and U. Alon, Science **305**, 1107 (2004).

[58] Y. Artzy-Randrup and L. Stone, Phys. Rev. E **72**, 056708 (2005).

[59] C. I. Del Genio, H. Kim, Z. Toroczkai, and K. E. Bassler, PLoS ONE **5**, e10012 (2010).

[60] J. Blitzstein and P. Diaconis, Internet Mathematics **6**, 489 (2010).

[61] V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, Phys. Rev. Lett. **94**, 018102 (2005).

[62] M. P. van den Heuvel, C. J. Stam, M. Boersma, and H. E. Hulshoff Pol, NeuroImage **43**, 528 (2008).

[63] S. Hayasaka and P. J. Laurienti, NeuroImage **50**, 499 (2010).

[64] P. Fransson, U. Åden, M. Blennow, and H. Lagercrantz, Cereb. Cortex **21**, 145 (2011).

[65] L. Tian, J. Wang, C. Yan, and Y. He, NeuroImage **54**, 191 (2011).

[66] M. Guye, G. Bettus, F. Bartolomei, and P. J. Cozzone, Magn. Reson. Mater. Phy. **23**, 409 (2010).

[67] M. Small, D. Yu, and R. G. Harrison, Phys. Rev. Lett. **87**, 188101 (2001).

[68] M. Breakspear, M. Brammer, and P. A. Robinson, Physica D **182**, 1 (2003).

[69] T. Nakamura and M. Small, Phys. Rev. E **72**, 056216 (2005).

[70] C. J. Keylock, Phys. Rev. E **73**, 036707 (2006).

[71] T. Suzuki, T. Ikeguchi, and M. Suzuki, Physica D **231**, 108 (2007).

[72] M. C. Romano, M. Thiel, J. Kurths, K. Mergenthaler, and R. Engbert, Chaos **19**, 015108 (2009).

[73] W. Hoeffding and H. Robbins, Duke Math. J. **15**, 773 (1948).

[74] E. Slutsky, Metron **5**, 3 (1925).